

**MEMORIA  
PROYECTO USISA  
SISTEMA DE RECOMENDACIÓN DE PRODUCTOS**

**Andrés Membrillo Pérez**

## **1-Introducción**

Esta memoria recopila los pasos que he llevado a cabo en el proyecto realizado para la empresa Unión Salazonera Isleña S.A., nombrada a partir de ahora como USISA.

Ubicada en Isla Cristina (Huelva), USISA ha liderado el sector alimentario, consolidándose como la conservera de pescado más grande en Andalucía.

Verticalmente integrada, USISA abarca todas las actividades, desde la producción hasta el servicio postventa.

Sergio Baeza-Herrazti, en calidad de representante de USISA, está en busca de un sistema de recomendación de productos destinado a la tienda online [www.usisa.com](http://www.usisa.com). El objetivo principal de este sistema es proporcionar recomendaciones personalizadas a los usuarios antes de finalizar sus compras.

## **2- Datos**

La empresa nos proporcionó todos los datos existentes de la tienda online. Aquí nos encontramos el primer obstáculo: en 2022 la empresa cambió de sistema de base de datos, y por lo tanto en la actualidad existen datos registrados en dos bases de datos diferentes:

### **Base de datos antigua:**

Formato archivo: `xlsx`

Periodo: Julio 2017 - marzo 2022

Shape: 10769 filas× 7 columnas

Registra por fila cada producto del carrito de compra, por ejemplo si un cliente compra 5 productos distintos, el sistema registra 5 filas con el mismo `id_order`.

Columnas:

`'id_order'`, `'Cliente'`, `'fecha'`, `'Referencia del pedido '`, `'product_name'`,  
`'product_quantity'`, `'total_price_tax_incl'`.

## Base de datos nueva:

Formato archivo: csv

Periodo: Marzo 2022 - **noviembre 2023**

Shape: 2019 filas × 95 columnas

Registra todos los productos del carrito de compra, en una misma fila.

Columnas:

'order\_id', 'order\_number', 'order\_date', 'paid\_date', 'status', 'shipping\_total',  
'shipping\_tax\_total', 'fee\_total', 'fee\_tax\_total', 'tax\_total', 'cart\_discount',  
'order\_discount', 'discount\_total', 'order\_total', 'order\_subtotal', 'order\_key',  
'order\_currency', 'payment\_method', 'payment\_method\_title', 'transaction\_id',  
'customer\_ip\_address', 'customer\_user\_agent', 'shipping\_method', 'customer\_id',  
'customer\_user', 'customer\_email', 'billing\_first\_name', 'billing\_last\_name',  
'billing\_company', 'billing\_email', 'billing\_phone', 'billing\_address\_1',  
'billing\_address\_2', 'billing\_postcode', 'billing\_city', 'billing\_state', 'billing\_country',  
'shipping\_first\_name', 'shipping\_last\_name', 'shipping\_company', 'shipping\_phone',  
'shipping\_address\_1', 'shipping\_address\_2', 'shipping\_postcode', 'shipping\_city',  
'shipping\_state', 'shipping\_country', 'customer\_note', 'wt\_import\_key',  
'shipping\_items', 'fee\_items', 'tax\_items', 'coupon\_items', 'refund\_items',  
'order\_notes', 'download\_permissions', 'meta:\_wcpdf\_invoice\_number',  
'meta:\_wcpdf\_invoice\_date', 'meta:\_wcpdf\_invoice\_number\_data',  
'meta:\_wcpdf\_invoice\_date\_formatted', 'meta:\_wcpdf\_invoice\_settings',  
'meta:\_ppcp\_paypal\_fees', 'line\_item\_1', 'line\_item\_2', 'line\_item\_3', 'line\_item\_4',  
'line\_item\_5', 'line\_item\_6', 'line\_item\_7', 'line\_item\_8', 'line\_item\_9', 'line\_item\_10',  
'line\_item\_11', 'line\_item\_12', 'line\_item\_13', 'line\_item\_14', 'line\_item\_15',  
'line\_item\_16', 'line\_item\_17', 'line\_item\_18', 'line\_item\_19', 'line\_item\_20',  
'line\_item\_21', 'line\_item\_22', 'line\_item\_23', 'line\_item\_24', 'line\_item\_25',  
'line\_item\_26', 'line\_item\_27', 'line\_item\_28', 'line\_item\_29', 'line\_item\_30',  
'line\_item\_31', 'line\_item\_32', 'line\_item\_33'.

### **3-Objetivo**

El propósito central de este proyecto reside en desarrollar un sistema de recomendación para la tienda en línea de USISA. Con los datos existentes se pueden obtener las siguientes variables:

Identificador de pedido: gracias al identificador del pedido podemos agrupar todos los productos comprados en cada pedido.

Identificador de producto: hay 110 productos registrados con distinto identificador.

La esencia del sistema de recomendación se fundamenta en la capacidad de aprovechar la información derivada de las compras anteriores. Este sistema tiene como objetivo principal prever y sugerir productos a los usuarios, basándose en la similitud de sus patrones de compra con otros usuarios. La premisa fundamental es analizar y comprender las elecciones y preferencias pasadas de los usuarios para anticipar sus necesidades futuras, proporcionándoles recomendaciones personalizadas que mejoren significativamente su experiencia de compra en la plataforma. La similitud en los patrones de compra entre usuarios se convierte así en la piedra angular para ofrecer sugerencias relevantes y adaptadas a cada individuo, impulsando así la eficacia y la utilidad del sistema de recomendación en la tienda online.

### **4-Limpieza y manipulación de datos**

En el contexto del sistema de recomendación hemos tomado la decisión de sacrificar volumen de datos y utilizar exclusivamente la base de datos más reciente (df\_nuevo). La razón principal de esta elección radica en que en esta nueva base de datos, los productos están registrados e identificados de manera correcta y consistente. Por el contrario, no hemos podido incorporar los datos de la base de datos antigua debido a la falta de identificadores para los productos; además, observamos que un mismo producto está registrado con distintos nombres, lo cual complica la integridad y coherencia de los datos.

Actualmente no tenemos datos suficientes del dataframe nuevo, por lo que es fundamental contar con todos los datos históricos de la tienda online. De esta manera hemos fusionado el df\_nuevo con el df\_antiguo para tener una visión completa de la evolución a lo largo del tiempo.

Primero hemos limpiado el **df\_nuevo**:

Eliminación de columnas.

Estandarizado de variables.

Transformación de fecha a formato date time.

Selección de los pedidos con status completado.

Eliminación de pedidos de prueba.

Eliminación de NaNs.

Renombrar columnas.

Columnas df\_nuevo después de limpieza:

*'fecha', 'order\_id', 'cliente\_nombre', 'producto\_nombre', 'producto\_id',  
'product\_cantidad', 'total\_precio'.*

Para el sistema de recomendación solo necesitamos las columnas identificadoras de pedido y de producto, pero también hemos generado un data frame con el nombre `ventas_por_producto` para el EDA, este conjunto de datos mide la cantidad de ventas totales de cada producto.

## 5-Análisis exploratorio de los datos

El Análisis Exploratorio de Datos (EDA) abarca gráficos que cuantifican las ventas totales de cada producto registradas en la base de datos más reciente de la tienda en línea, es decir, a partir del año 2022.

Hemos creado cinco gráficos de barras que establecen un ranking de los 110 productos disponibles para la venta, clasificándolos desde el más vendido hasta el menos vendido.

Además, hemos incluido un box plot que visualiza la distribución de las cantidades vendidas por cada producto.

## 6- Sistema de Recomendación.

### 6.1-Preprocesamiento de datos

Se utiliza TransactionEncoder para convertir la lista de productos en una matriz binaria (one-hot encoding) y se realiza un train\_test\_split para dividir los datos en conjuntos de entrenamiento y prueba.

### 6.2-Algorithm Apriori

- Objetivo: El algoritmo Apriori se utiliza para descubrir conjuntos de productos que aparecen con frecuencia juntos en los pedidos.
- Implementación: Se aplica el algoritmo Apriori al conjunto de entrenamiento (train\_basket) utilizando la función apriori. Se establece un umbral de soporte mínimo del 0.5%, lo que significa que sólo se considerarán conjuntos de productos que aparezcan en al menos el 0.5% de los pedidos.
- Resultado: Se obtienen conjuntos frecuentes de productos y sus niveles de soporte.

### 6.3-Reglas de asociación

- Objetivo: A partir de los conjuntos frecuentes encontrados con Apriori, se generan reglas de asociación para capturar patrones de compra.
- Implementación: Se utiliza la función association\_rules para generar reglas de asociación a partir de los conjuntos frecuentes. Se establece un umbral de confianza del 60%, lo que significa que sólo se considerarán las reglas que tengan al menos un 60% de confianza.
- Información extraída: Cada regla de asociación consta de un conjunto de productos antecedentes y consecuentes, junto con medidas como confianza, soporte y lift. La confianza representa la probabilidad condicional de que los productos consecuentes se compren dado que se han comprado los antecedentes.
- Resultado: Se obtiene un conjunto de reglas de asociación que representan patrones de compra significativos, donde la confianza es un indicador clave de la fuerza de la asociación entre los productos.

Estos dos pasos son cruciales para la generación de recomendaciones, ya que el algoritmo Apriori y las reglas de asociación permiten identificar patrones de compra comunes. Estos patrones se utilizan luego para sugerir productos adicionales que tienen una alta probabilidad de ser comprados junto con los productos de un pedido dado, contribuyendo así a un sistema de recomendación efectivo.

#### **6.4-Generación de recomendaciones para el conjunto de prueba y para un nuevo pedido**

Se crean funciones para obtener recomendaciones de productos en el conjunto de prueba, filtrándose según reglas de asociación. El resultado se presenta como un diccionario, donde las claves son los índices de los pedidos en prueba y los valores contienen recomendaciones de productos con sus nombres asociados.

También se muestra un ejemplo de cómo obtener recomendaciones para un nuevo pedido específico.

#### **7-Conclusión**

En el desarrollo de este proyecto, hemos logrado implementar un robusto sistema de recomendación para la tienda online de USISA, utilizando datos de compras pasadas y reglas de asociación. Sin embargo, identificamos una oportunidad clave para mejorar aún más la precisión y personalización de nuestras recomendaciones. La incorporación de ratings específicos para cada producto permitiría usar otros modelos de recomendación y ofrecer sugerencias más ajustadas a las preferencias individuales de los usuarios, enriqueciendo significativamente la experiencia de compra.

En vista de este hallazgo, se recomienda encarecidamente la inclusión del registro de ratings para cada producto en la base de datos. Este enfoque proporciona una fuente invaluable de información adicional, potenciando la capacidad predictiva del sistema de recomendación y consolidando aún más la posición de la tienda online de USISA como un destino de compras altamente personalizado y satisfactorio para sus clientes.